Modelling Subordinate Conjunctions in STAG: A Discourse Perspective

Timothée Bernard

Modelling Subordinate Conjunctions in STAG: A Discourse Perspective

Timothée Bernard

Université Paris Diderot - ALPAGE
timothee.bernard@inria.fr

Abstract Among the discourse connectives – lexical items conveying discourse relations – are the subordinate conjunctions (SubConjs), like *because*, *even if* or *although*. SubConjs have generally been considered a homogeneous category, however previous work has shown they can be divided into two classes according to their syntactic and semantic properties. Similarly, attitude verbs and reporting verbs (AVs) have two different uses in discourse: *evidential* and *intentional*. Drawing from these observations, we propose a STAG model of SubConjs and AVs taking into account both their syntactic and discursive properties.

Keywords: discourse, STAG, subordinate conjunctions, syntax, semantics

1 Introduction

At the discourse level, sentences and propositions are related by *discourse relations* (DRs). DRs can either be *implicit*, i.e semantically inferred, or *explicit*, i.e. lexically signalled. The most common markers of explicit DRs are *discourse connectives*, a group mainly composed of conjunctions, prepositions and adverbs. For instance, (1a) involves an implicit *Consequence* relation and (1b) a *Concession* one explicitly signalled by the *but* connective. Following the conventions of the Penn Discourse TreeBank (PDTB, [13]), we refer to the two arguments of DRs as Arg_1 and Arg_2 and use italics and bold face respectively to indicate the spans of text for each argument (when such spans of text appears) while the connective lexicalising the relation, if any, is underlined.

(1) a. *Fred was sick.* **He stayed at home**.
b. *Fred was sick.* <u>But</u> **he came to work**.

We are working with Synchronous Tree Adjoining Grammar (STAG, [15]), a formalism providing a way to describe both syntax and semantics simultaneously, making explicit how they relate with each other. To our knowledge, not much attention has been paid to modelling subordinate conjunctions (SubConjs) from a discourse point of view in STAG. D-STAG [3] analyses discourse with STAG structures, and thus models discourse connectives, but does not take into account the specificities of SubConjs that are discussed here. This is however a necessary step toward both operational discourse parsers and convincing discourse generation systems, and also the purpose of this mainly theoretical work.

The paper is organised as follows. Section 2 presents relevant work highlighting the aspects of SubConjs that we aim to model. In Section 3, we use linguistic tests to determine more precisely the interactions of SubConjs with diverse scope operators. This leads to our STAG proposition presented in Section 4. Section 5 concludes the paper.

2 Relevant Work on Subordinate Conjunctions

2.1 Non-Alignment of Syntactic and Discourse Arguments

It has been shown by a number of works (see [5] for English and [4] for French) that the propositional content of a (syntactic) argument of a discourse connective is not always a (semantic/discourse) argument of the DR lexicalised by the connective. Such mismatches often arise with *attitude verbs* (*to think*, *to know*, etc.) and *reporting verbs* (*to say*, *to deny*, etc.), both grouped here under the label 'AV'. When an AV together with the clause it introduces is an argument of a discourse connective, the AV may (2a) or may not (2b) be included in the discourse argument of the corresponding DR.[1] Following [1], we say the AV is *intentional* in the first case and *evidential* in the second.

(2) a. *Fred went to Peru* although **Sabine thinks he never left Europe**.
b. *Fred went to Peru* although Sabine thinks **he did not go to Lima**.

It is interesting to note that contrarily to *although*, not all discourse connectives can be found with such non-alignments of the syntactic and discourse arguments. It is the case, for instance, of *because*, as illustrated in (3). [9], using the DR hierarchy of the PDTB, observes that a connective lexicalising a COMPARISON or an EXPANSION relation can often be found with a mismatch, whereas it seems impossible for a connective lexicalising a TEMPORAL or a CONTINGENCY relation.

(3) a. *Fred could not come* because **he was not in town**.
b. #Fred could not come because Sabine thinks he was not in town.

2.2 Two Types of Adverbial Clauses

A distinction between two types of adverbial clauses is made by [8]. The first type is the *central adverbial clause* (CAC), which adds an information (time, place, etc.) about the eventuality described in the matrix clause as in (4). The second type is the *peripheral adverbial clause* (PAC), whose function is to structure the discourse (expressing a concession, providing background information, etc.) as in (5).

[1] Why the AV is included or not in Arg_2 is discussed in [5] and [4]. One element is that the AV can be felicitously removed from (2b) while it cannot from (2a). Similarly, an attributing phrase such as *according to Sabine* can be substituted for the AV (with no change in meaning) only in (2b) and not in (2a).

(4) a. *Fred went to Brazil* while **he was a student**.
b. If **it is sunny**, *I'll go outside.*

(5) a. *Fred has been to Brazil* whereas **Sabine has never left Europe**.
b. If **it is sunny**, *why aren't you playing outside?*

Several phenomena are studied in [8] – coordination, ellipsis, ambiguity, and others related to scope, prosody, typography, etc. They all tend to show a greater integration of CACs into their matrix clause than PACs. We will expand on these observations concerning scope phenomena in the next section. It suffices for now to point out that negation and interrogation may scope over CACs but not over PACs. It should also be noted that a CAC cannot contain an epistemic modal if it is *speaker-oriented* (as in (6a) but not in (6c) where *may* is mainly 'John-oriented'), while a PAC can (see 6b). Expressed with the terms of [9]: the syntactic and discourse arguments of a conjunction must be aligned in the case of a CAC, while there can be a mismatch with PACs.

(6) a. #Mary accepted the invitation without hesitation after John may have accepted it. (from [8])
b. *The ferry will be fairly cheap*, while/whereas **the plane** may/will probably **be too expensive**. (from [8])
c. *John is worried* because **he may be ill**.

3 Projection Tests Applied to Subordinate Conjunctions

In order to model SubConjs, we need to understand how they semantically relate to the other components of the sentence. We therefore study them in the context of the five following patterns[2], related to the scope of diverse operators with respect to discourse connectives and their arguments:

Negation: It is not the case that A.
Conditional: If A, B.
Epistemic: It is possible that A.
Interrogation: A?
AV: Sabine thinks that A.

In these patterns, we replace A with 'A_1 CONJ A_2', where CONJ is a SubConj lexicalising a DR R, and try to figure out if Arg_1, Arg_2 and $R(\text{Arg}_1, \text{Arg}_2)$ are logically implied by the resulting sentence. Note that we do not constrain the syntactic structure of the sentence; we do not know *a priori* whether A is made up of a unique constituent or of multiple constituents diversely attached to the rest of the sentence. In this paper, we illustrate the results for *because* and *although* – introducing a CAC and a PAC respectively – although the examples can be extended to many other SubConjs.

[2] These patterns are commonly used to test projection properties [2].

Because (Explanation): Sentences in (7) are the result of applying the negation and the interrogation patterns to an instance of *because* lexicalising *Explanation*.

(7) a. It is not the case that *Fred was absent* because **he was sick**.
b. Was Fred absent because he was sick?

Interpret (7a) in the context of Fred's workplace. A local interpretation of the negation in the matrix clause (scoping only over *Fred was absent*) would be logically incoherent (in the sense that while it is semantically well-formed, it seems impossible or at least very hard to find a situation in which it would be true), so it must have a global interpretation. This is compatible with [8], as *because* specifies some aspect of the event in the matrix clause and thus introduces a CAC. But what do possible continuations tell us about the semantics of (7a)?[3] All sentences in (8) are possible and describe different situations:[4]

- with (8a), neither Arg_1 (*Fred is absent*), nor Arg_2 (*Fred is sick*), nor R (*Explanation*) are true;
- with (8b), only Arg_1 is true;
- with (8c), only Arg_2 is true;
- with (8d), both Arg_1 and Arg_2 are true, but not R.

(8) a. He was there and in perfect shape.
b. He was fine but he missed his train.
c. He still came, even if indisposed.
d. He was indeed ill, but he would have come anyway hadn't he had to take delivery of some important package.

The negation has therefore a global scope over $\text{Arg}_1 \wedge \text{Arg}_2 \wedge R(\text{Arg}_1, \text{Arg}_2)$ and none of these elements are semantically implied by the use of the *because*. We can notice that all these sentences are also acceptable answers for question (7b) when preceded by the negative *no*. So interrogation has the same properties as negation in terms of these conclusions.

In a similar way, the conditional pattern generates (9), which can be coherently followed by any sentence in (10). They illustrate the same four previous configurations: neither Arg_1 nor Arg_2 nor R are true with (10a), only Arg_1 is true with (10b), only Arg_2 with (10c), Arg_1 and Arg_2 true but not R with (10d).

(9) If *Fred got offended* because **Sabine teased him**, then it would mean that he is secretly in love with her.

[3] We accept any kind of continuation, including dialogue, as long as no correction DR nor 'Hey, wait a minute'-style device [14] is involved.

[4] It could be argued that (8a) and (8c) are no acceptable continuations of (7a), implicitly saying that the use of *because* presupposes the truth of Arg_1. However, this interpretation does not seem to be shared by all English speakers, and as it may involve a specific treatment of presupposition, we have preferred to leave it for future research.

(10) a. However, Fred is in very good mood and I know Sabine, she never teases anyone.
b. However, I don't think she was teasing him.
c. However, he didn't seemed annoyed at all.
d. However, I don't think that this is actually the reason.

The same conclusions can be drawn from all other patterns as well: the corresponding operators can have global scope over the whole 'A_1 because A_2' span.

Although (Concession): There has been a lot of discussion since Frege [6] about the semantics of *although*. According to [12], the *Concession* relation is not *at-issue* (which roughly means that while you can *express* it, you cannot *talk about* it; in particular it cannot be easily negated). It is also interesting to remark that this *Concession* is *speaker-oriented*: *although* cannot be used without the speaker committing herself to the relation, even if the connective is under the scope of AVs, which are *presupposition plugs*. Therefore, *although* is often cited as a *conventional implicature* trigger since [7].

First, let us notice that the negation pattern cannot be directly used with a coherent 'A_1 although A_2'.[5] To produce a satisfactory utterance such as (11c), it is necessary for the negation to be included in Arg_1. The impossibility of sentences such as (11b) shows that with *although* – contrary to what we have just seen with *because* –, a negation in the matrix clause always has a local scope. This is consistent with the analysis in [8], as *although* is not used to precise an event but to give some context for its interpretation and thus introduces a PAC.

(11) a. *Fred ate meat the other day* although **he is a vegetarian**.
b. #It is not the case that Fred ate meat the other day although he is a vegetarian.
c. *It is not the case that Fred refused to eat meat the other day* although **he is a vegetarian**.

Let's consider (12), from the interrogative pattern. Whereas (13a) is a perfectly acceptable answer to it, (13b) is not.[6] This tends to show that with *although*, Arg_2 and R are not at-issue and that the interrogation only concerns the content of the matrix clause.

(12) Did he eat meat although he is a vegetarian?

[5] The examples in (11) would be more compelling if *although* was replaced with *despite the fact that*. However, the lack of appropriateness of *although* comes from subtle differences in semantics and usage which are unrelated to the problem at stake. It is for the sake of simplicity and homogeneity that we have chosen to stick to *although*.

[6] It would be possible to continue (12) with *He is not a vegetarian anymore*, but this is more of a remark than an answer: the dialogue could continue with *You haven't answered my question*. Also note the use of *anymore*, which marks a revision.

(13) a. No, he refused.
b. #No, he is not a vegetarian anymore.

Yet, saying that the Arg_2 of a *Concession* is never at-issue would be taking shortcuts. It seems for example that (14a), from the AV pattern, can be felicitously followed by (14b) although it negates Arg_2. So in such a case, *he was sick* is under the semantic scope of *Sabine thinks*.

(14) a. Sabine thinks *Fred came to work* although **he was sick**.
b. But she is wrong, he had recovered several days ago.

Out of context, (14a) seems intuitively to imply that Fred was actually sick; this is a default reading. The utterance is ambiguous: the Arg_2 may or may not be under the semantic scope of the AV, the latter being the default interpretation.

Summary: Tab. 1 summaries these properties and those of two other SubConjs, *after* and *whereas*, which could not be discussed here due to lack of space. *although*, *whereas* and other SubConjs are ambiguous between *Contrast* and *Concession* (at least, [13]), but they have the same properties as long as they introduce a PAC.

It seems that conjunctions introducing a PAC ('peripheral conjunctions', PCs) all share the same behaviour; they allow mismatches for Arg_2, the speaker is always committed to the relation conveyed and in the (very probable) default reading the speaker is also committed to Arg_2. Conjunctions introducing a CAC ('central conjunctions', CCs) also share some properties; they do not allow any mismatch for Arg_2, the commitment of the speaker toward the relation conveyed is always subject to the modifiers used in the patterns. The status of Arg_2, however, depends on the conjunction.

R (type)	CONJ	Arg_2	$R(\text{Arg}_1, \text{Arg}_2)$	mismatch for Arg_2
Explanation (central)	because	−	−	−
Narration (central)	after	(+)	−	−
Concession (peripheral)	although	(+)	+	+
Contrast (peripheral)	whereas	(+)	+	+

Table 1. The Arg_2 and $R(\text{Arg}_1, \text{Arg}_2)$ columns show if the truth of these propositions are still implied by the use of CONJ in the studied patterns: '−' means 'no', '+' means 'yes' and '(+)' means 'yes in the default reading'; '+'/'−' in the last column indicate whether CONJ can or cannot be found with a mismatch concerning its Arg_2. Arg_1 is always subject to the operators used in the various patterns.

4 Our Proposition in STAG

We now turn to STAG and propose a basic model for AVs, CCs and PCs that reflects the properties observed in the previous sections. But before that, let us explain what this formalism is and how it works.

4.1 TAG and STAG

The Tree Adjoining Grammar formalism (TAG), on which is based STAG, was introduced in [10]. In TAG, words are represented as tree structures of two kinds. On the one hand are the *initial trees* (named with α), whose interior nodes are labelled with non-terminal symbols and whose leaves are either labelled with a terminal symbol, either labelled with a non-terminal symbol and marked with $\downarrow$. In the last case, the leaf is said to be a *substitution site*. On the other hand are the *auxiliary trees* (named with β), which are similar to initial trees except that they have (exactly) one leaf, the *foot node*, that is labelled with the same non-terminal as the root and is marked with $*$ instead of $\downarrow$.

These trees are meant to combine into sentences using two operations: *substitution* and *adjunction*. A set of such operations is called a *derivation tree* and the resulting tree is called a *derived tree*. A substitution consists in replacing a substitution site with an whole initial tree whose root must be labelled with the same symbol that the substitution site. An adjunction consists in inserting at some interior node an auxiliary tree whose root must also be labelled with the same symbol that the target node. Both operations are illustrated in the upper part of Fig. 1 with a syntactic grammar: a substitution and an adjunction are represented on the left side while the resulting tree appears on the right side.

The idea of STAG [15] is to pair two TAGs together to perform parallel operations. Thus, in STAG, a lexical entry is a pair of TAG trees with a set of links precising the coupling between the two. A link is a pair of nodes, one from each tree, here marked as $^{[1]}$, $^{[2]}$, etc. Only on a linked node can an adjunction or a substitution be performed. When a substitution (resp. an adjunction) occurs at a node, the parallel substitution (resp. adjunction) must also occur on the other node of the link. This principle is illustrated in Fig. 1 with the coupling of a syntactic grammar (top) and a semantic one (bottom).

Note that we allow multiple adjunctions on the same node – up to one for each link on that node. In such a case, the order of the adjunctions must be specified in order to describe the resulting derived tree pair. Otherwise, the derivation tree is *underspecified* and is used to represent all the derived trees corresponding to all possible orders.

4.2 Lexical Entries

AVs: To take into account the two *evidential* and *intentional* uses of AVs, we propose an initial TAG pair for these verbs in addition to the auxiliary one traditionally used (Fig. 2). Auxiliary trees for AVs are motivated by long distance extractions [10], where *John says* is equivalent to the adjunction of *according to John*. But this equivalence does not hold for intentional AVs in a discourse structure. Furthermore, adjunction is generally used to indicate semantic modifiers, whereas an intentional AV provides the main predicate that is argument of a DR and does not merely indicate attribution. That is why we think it makes sense to model them with initial trees rather than auxiliary trees.

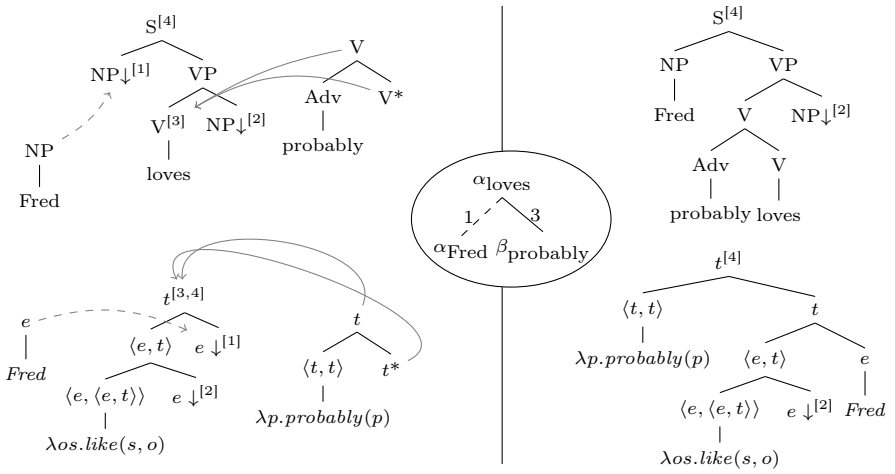

Figure 1. Substitution of α_{Fred} at link [1] in α_{loves} and adjunction of β_{probably} at link [3]; circled in the middle is the derivation tree; the derived trees are on the right.

In our model their semantics is also slightly different: evidential AVs use predicates that are 'erased' when in a DR. This is achieved by introducing rewriting rules of the form $Contrast(p, think(a, q)) \rightarrow Contrast(p, q)$. Conversely, unnatural mismatches can be avoided by discarding any analysis displaying an evidential AV predicate as argument of a central DR: $Explanation(p, think(a, q)) \rightarrow \bot$.

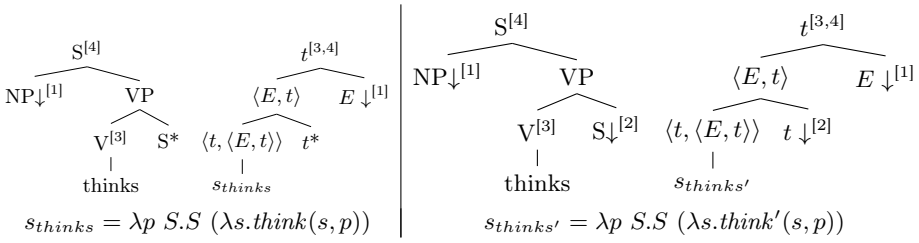

Figure 2. AVs: β_{thinks} (evidential) and α_{thinks} (intentional)

Subordinate conjuctions: Similarly, the difference in syntax and semantics between CACs and PACs can be explained with different structures for CCs and PCs as in Fig. 3.[7] The most significant aspect of these structures is that CCs are auxiliary trees whereas PCs are initial trees.

[7] The presence of the SBAR node for CCs is necessary because of the possibility of cleft sentences (*It is because A that B*), which shows that there exists such a constituent. No cleft sentences are observed with PCs.

Indeed, consider a CC adjoined to its matrix clause. If a semantic modifier (a negation, for instance) is also adjoined to this clause, depending on where this adjunction is done relatively to the adjunction of the CC, this modifier may or may not scope over the DR. With a PC, however, because the matrix clause is substituted into the connective, any modifier is necessarily dominated by the connective and thus only the local scope is possible.

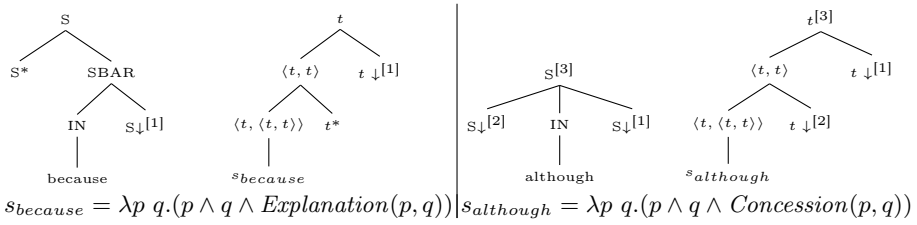

Figure 3. SubConjs: β_{because} (CC) and α_{although} (PC)

Sentence structures: [11] proposes that in sentence structures, verbal modifiers are adjoined in the semantic tree at a lower node than AVs. Doing so avoids (unnatural) interpretations of the former scoping over the latter. However, as seen in the previous section, while CCs are sentence modifiers like AVs they do present scope ambiguity when confronted with verbal modifiers such as negation. This is why, as illustrated in Fig. 4, we consider adding to sentence structures another adjunction site on the S-node (link [3]) whose semantic counterpart is at the same node as verbal modifiers' one. We can use features to restrict the other S-site (link [2]) to AVs and conversely to force them to adjoin there.

Fig. 4 also shows the derivation trees obtained when a negation (or any other verbal modifier) is present in the matrix clause of a SubConj. As expected, the negation can either have a local or a global scope with a CC while it always has local scope with a PC.

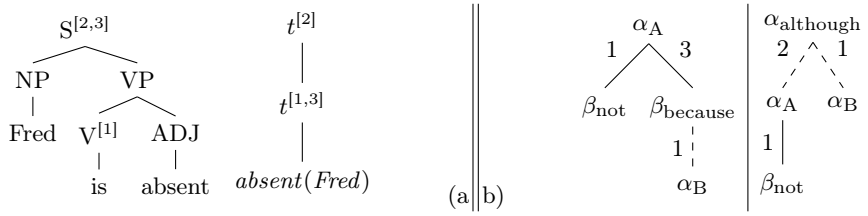

Figure 4. A typical sentence structure (a), accompanied by the derivation trees for *not A CONJ B* with a CC (b. left) or a PC (b. right). Because in α_{A} links [1] and [3] are at the same semantic node, the left tree is a scope neutral representation yielding one syntactic tree but two semantic ones depending on the order of the adjunctions.

5 Discussion and Perspectives

Tab. 2 shows the derivation trees for sentences of the form *A CONJ Sabine thinks B* as in (2). While the rewriting rules we have introduced discard the use of an evidential AV with a CC, there is an ambiguity for PC that can only be resolved using the particular semantics of the lexicalised relations.

Tab. 3 shows the derivation trees for sentences of the form *Sabine thinks A CONJ B* as in (14a). In this configuration, it is possible for the AV to scope over the relation. Its evidential or intentional status is then undetermined (it would depend on another discourse relation) and we have arbitrarily chosen to represent it with the traditional auxiliary structure in the 'External AV' column. As previously, the various ambiguities are natural ones and can only be resolved using world knowledge.

These analyses generate the expected derived trees, on both the syntactic and the semantic side (though we lack space to exhibit them here). However, our model should be refined: it does not yet account for the projection of peripheral DRs as observed in (14b) nor for the default projection of Arg_2 for some CCs such as *after* (not discussed here, but that some also see in *because* - see note 4).

Furthermore, discussing the interaction between AVs and SubConjs, we have said that evidential AVs could be replaced with attributing prepositional phrases such as *according to Sabine* (15a). The other way around is not true, because such expressions can be found within CACs (15b), where evidential are forbidden. In fact, it seems that in these cases the attribution not only scopes over Arg_2 but also over the relation itself. Attributing prepositional phrases therefore exhibit very interesting behaviours at the discourse level and will likely prove challenging to model considering their relatively free position in the sentence (15).

We also intend on extending this study to other connectives (especially adverbials, such as *instead* and *otherwise*) with the ultimate goal of building a parser capable of providing analyses coherent at the syntactic, semantic and discourse levels.

(15) a. *Fred could not come* even though, according to Sabine, **he was really looking forward to it**.
b. *Fred could not come* because, according to Sabine, **he was not in town**.
c. *Fred could not come* even though **he was**, according to Sabine, **really looking forward to it**.
d. *Fred could not come* even though **he was really looking forward to it**, according to Sabine.

Acknowledgements

I would like to particularly thank my supervisor Laurence Danlos for her invaluable advice, as well as my friends Rachel and Emily for their help with the English language.

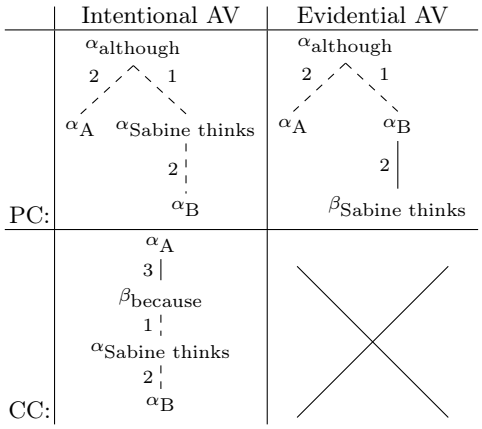

Table 2. Derivation trees for sentences of the form *A CONJ Sabine thinks B*.

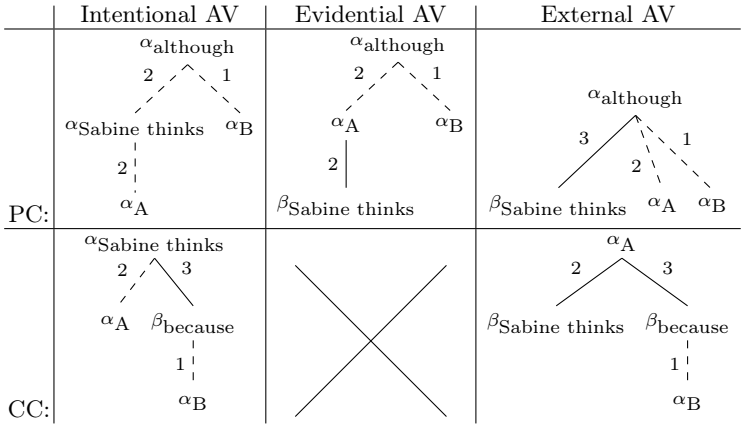

Table 3. Derivation trees for sentences of the form *Sabine thinks A CONJ B*.

References

1. N. Asher, J. Hunter, P. Denis, and B. Reese. Evidentiality and intensionality: Two uses of reportative constructions in discourse. In *Workshop on Constraints in Discourse Structure*, Maynooth, Ireland, 2006.
2. G. Chierchia and S. McConnell-Ginet. *Meaning and Grammar: An Introduction to Semantics*. MIT Press, 1990.
3. L. Danlos. D-STAG: a Formalism for Discourse Analysis based on SDRT and using Synchronous TAG. In P. de Groote, editor, *Proceedings of FG'09*. INRIA, 2009.
4. L. Danlos. Connecteurs de discours adverbiaux: Problèmes à l'interface syntaxe-sémantique. *Linguisticae Investigationes*, 36(2):261–275, December 2013.
5. N. Dinesh, A. Lee, E. Miltsakaki, R. Prasad, A. Joshi, and B. Webber. Attribution and the (Non-)Alignment of Syntactic and Discourse Arguments of Connectives. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, pages 29–36, Ann Arbor, Michigan, June 2005. ACL.
6. G. Frege. Sense and Reference. *The Philosophical Review*, 57(3):209–230, 1948.
7. H. Grice. Logic and conversation. In P. Cole and L. Jerry, editors, *Syntax and semantics 3: Speech acts*, pages 41–58. Academic Press, San Diego, CA, 1975.
8. L. Haegeman. The syntax of adverbial clauses and its consequences for topicalisation. In M. Coene, G. De Cuyper, and Y. D'Hulst, editors, *Current Studies in Comparative Romance Linguistics*, number 107 in APiL, pages 61–90. Antwerp University, 2004.
9. J. Hunter and L. Danlos. Because We Say So. In *Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language*, CAtoCL, pages 1–9, Gothenburg, Sweden, April 2014. ACL.
10. A. Joshi. An introduction to tree adjoining grammars. *Mathematics of language*, 1:87–115, 1987.
11. R. Nesson and S. Shieber. Simpler TAG semantics through synchronization. In *Proceedings of FG 2006*, pages 129–142, Malaga, Spain, 2006.
12. C. Potts. Presupposition and Implicature. In S. Lappin and C. Fox, editors, *The Handbook of Contemporary Semantic Theory*, pages 168–202. Wiley-Blackwell, 2 edition, 2015.
13. R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber. The Penn Discourse TreeBank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, Marrackech, Morocco, June 2008.
14. B. Shanon. On the Two Kinds of Presuppositions in Natural Language. *Foundations of Language*, 14(2):247–249, 1976.
15. S. Shieber and Y. Schabes. Synchronous Tree-adjoining Grammars. In *Proceedings of the 13th Conference on Computational Linguistics - Volume 3*, COLING '90, pages 253–258, Stroudsburg, PA, USA, 1990. ACL.